DEC 29 1987

COMPLETED

ED 1.310/2:277158

THE Northwest Regional Educational Laboratory

PROGRAM REPORT

ED277158

BEYOND THE WALL CHART:
APPRAISAL OF STATEWIDE EDUCATIONAL PROGRESS

A Policy Issues Paper

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X

Prepared for the Chief State School Officers
of the Northwest and Pacific

Kenneth H. Hansen
Principal Author

August 1986

Northwest Center for State Education Policy Studies

EA 019 091

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

Sponsored by OERI Office of Educational Research and Improvement U.S. Department of Education

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number 400-86-0006. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

DOCUMENT RESUME

ED 277 158 EA 019 091

AUTHOR Hansen, Kenneth H.
TITLE Beyond the Wall Chart: Appraisal of Statewide Educational Progress. A Policy Issues Paper. Program Report.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Northwest Center for State Educational Policy Studies.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Aug 86
CONTRACT 400-86-0006
NOTE 15p.; Prepared for the Chief State School Officers of the Northwest and Pacific.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; *Educational Policy; Elementary Secondary Education; Evaluation Criteria; Evaluation Needs; Measurement Objectives; Policy Formation; *State Programs

ABSTRACT

State educational policy-makers are showing increasing interest in using statewide educational assessment and evaluation programs (1) to evaluate student thinking skills and affective learning as well as traditional academic achievement, (2) to evaluate teachers at all stages of their careers, and (3) to assess the impact of school reform efforts. Whether actions to implement these evaluation program elements have been taken or whether they are currently being planned, there are policy issues involved that should be reviewed. One set of issues concerns identification of the purposes of the evaluation effort--an identification necessary if appropriate methods and instruments are to be selected. Establishing legitimate and informative criteria for making evaluative judgments is a second important policy concern. Policies relating to the appropriate extent of testing--clarifying how much is too much and how much not enough--are also required. The needs of governmental bodies for different information can lead to conflicts that may be resolved by adopting policies that recognize the legitimate interests of all agencies. The social implications of evaluation programs must also be considered and policies devised to control program effects. Effective policy-making can enhance the prospects for developing programs that provide data for significant planning. (PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEYOND THE WALL CHART:

APPRAISAL OF STATEWIDE EDUCATIONAL PROGRESS

Whatever strengths or weaknesses the "wall charts" produced by the U.S. Department of Education may possess, they have undeniably served to focus attention on statewide assessment and evaluation of educational progress. Cross-state comparisons, particularly of student achievement scores, have so far aroused the most popular interest, caused the most controversy, and dominated a great deal of the discussion about the validity and utility of the data presented in the charts.

Increasingly, however, interest and concern of state educational policy makers are expanding into areas which transcend the wall chart: continuing concern with measures of academic achievement, of course, but interest in other indices of educational progress as well. Three such concerns of a broader nature have emerged.

The first is an interest in assessing and evaluating a wider range of student learnings, not only the more traditional academic ones, but higher order thinking skills and affective learnings as well. The second is a concern with more demanding evaluation of teachers at every stage in their careers, from admission into teacher education programs through constant appraisal of on-the-job performance. The third is a determination to find ways to use assessment and evaluation results (and other data collection devices such as profiling) as a means of determining the impacts of the school reform movement. All of these are concerns well beyond the wall chart level.

Basic to the establishment and improvement of broadscale statewide assessment programs are a number of important policy issues--issues not of technical or programmatic details, essential as those are, but issues of the basic purpose and fundamental direction of the entire state-sponsored or state-operated program. One problem with attempting to consider basic policy questions is that it may appear to be too late in the game for such activity--the programs are already in place, and the pressures to keep them going are so strong as to preclude any significant changes at this point. It may well be, however, that it is for just this reason the programs should be re-examined from the standpoint of policy: many of them may have inadvertently been built more in response to pressure than as an expression of deliberately developed state educational policy. This is not at all to say that these programs were established thoughtlessly or whimsically, or that the underlying motivations were not sound. Rather, pressures may have simply taken precedence over policy considerations.

At the state level, for example, there has often been a very-well-intentioned push from gubernatorial or legislative quarters (ranging in intensity from exhortation through formal statutory mandate) to establish or expand a statewide program to "test the kids to see what they are learning," or to establish a statewide competency test for high school graduation or for teacher certification or recertification. These are probably worthy goals, but in putting into place such assessment and evaluation programs, there may not have been time to consider long-range educational implications.

At the national level, both federal and nonfederal agencies or organizations are making commendable efforts to improve the quality and depth of the education data which are being gathered and reported (most of them necessarily from the states), and which can be used to give a better picture of the strengths and weaknesses of the American educational system. The Department of Education (primarily through the Office of Educational Research and Improvement and the Center for Statistics, plus two nongovernmental but partially federally supported projects, the National Assessment of Educational Progress and the CCSSO Assessment Center) is engaged in activities which both reflect and affect what is being done in the states. With so many actors on the scene, it seems almost inevitable that national organization or agency needs and programs may not always reflect what individual states see as their own goals and priorities. At the very least, it would seem that systematic analysis, careful value judgements, and clear policy determinations which respond to these many pressures are needed at the state policy-making level.

If it can be generally agreed that we are now beyond the wall-chart stage in our formulation of statewide assessment and evaluation programs, that statewide programs now in place might benefit from some updated policy analysis, and that the variety of state-level and national forces seeking to influence state data-collection programs appear to be increasing--if all of this is true, it may be useful to examine some specific policy issues which bear directly on state assessment and evaluation programs/efforts. These issues--<u>purposes</u>, <u>criteria</u>, <u>balance</u>,

autonomy, and pervasive societal concerns--are important in all of the statewide evaluation and assessment programs, whether they are addressed to student achievement, teacher competency, or whole-system performance.

Purposes

Initial establishment of a statewide assessment program is, in itself, an expression of purpose and direction--an educational policy decision. As the program becomes an ongoing operation, or as changes and refinements are made in it, subtle changes may occur. The original purposes may have become blurred. Alternatively--or additionally--the original purposes and directions may have been faithfully adhered to, even though changing conditions and needs would suggest the desirability of re-examination of the earlier goals and directions. There may be need for the education decision makers to ask again, "Just why are we doing this?"

It is possible, for example, that a state assessment program was initiated because it was either mandated by statute or required as a response to public pressure--certainly legitimate and sufficient reasons for starting a program, but not necessarily ones for continuing it. More compelling reasons are needed.

What should be the purposes of a state assessment program can be determined only by those responsible for the state educational system--ideally not the legislature or the state board of education making unilaterial decisions, but through the cooperative efforts of all of those with a legitimate stake in the system. It is not particularly crucial that a given purpose or set of purposes be chosen; what is important is that the choice be deliberately and consciously made following an examination of a number of relevant options.

If the central purpose of a student-testing program is basically a statewide assessment of academic learnings at the various grade or age levels, similar to the NAEP program, one cluster of instruments and sampling techniques will be appropriate, and one kind of data will become available. If the purpose of the program is to elicit information needed for making determinations about an individual's educational performance and instructional needs, a wholly different set of instruments will be needed, with the focus doubtless shifting from a state-administered to a locally-administered program. But if the individual results are to be used in determining eligibility for grade promotion or high school graduation, under state-mandated standards, the whole focus of the program again shifts to the state.

If the established purpose of the program goes beyond an assessment of academic progress to include elements of personal growth and social development, different instruments, scales, and procedures will be called for, with the state again taking the initiative and giving direction, but with the program being primarily under the aegis of the local district.

On the other hand, if the state-determined emphasis of the assessment program is directed toward appraisal of the entire educational enterprise, rather than toward assessing individual students or classes of students, the focus must expand to include curriculum, instruction, learning materials, school climate, quality of educational leadership, and other factors affecting the outcomes of education. Specifically, if we want to find out whether the "educational reform" or the "school improvement" efforts have made a real difference, it would be necessary for the statewide assessment and evaluation program to be expanded well beyond traditional achievement testing.

With so many possible and plausible kinds of purposes and directions which might be established for a statewide assessment and evaluation program, it certainly would appear to be prudent and productive to return frequently to a consideration of the basic policy question: "Why are we doing this?"

Criteria

Inescapably related to the basic policy questions of purpose and direction is the question of the criteria to be used in making judgements both about individual educational progress and overall system improvement. There are several policy issues involved here.

One of these issues involves the selection and validation of the "indicators" which will be used to judge the effectiveness of the educational system. All sorts of questions complicate the issue. How can the so-called "input" indicators be legitimately correlated to "outputs?" What indicators have commonality enough to permit appropriate comparisons between and among individual schools, school districts, and states? How many indicators need to be used--not just the "nice to know" ones, but the essential ones? Fortunately, a great deal of thoughtful consideration and technical analysis of indicators now underway should give policy makers some helpful guidance in answering these and related questions.

In addition to the somewhat formal "indicators" which can be used to measure educational progress and to make at least a goodly number of valid comparisons, other more subtle--perhaps essentially philosophical--considerations enter into the deliberations of policy makers concerned with the statewide assessment program. What relative importance should be attached to academic performance and to personal/social growth? Is

the number of dropouts--even if determined and reported far more precisely than is now the case--an important measure of the system's success unless we know more about the stay-ins who may really be no better off than those who have dropped out? What details of family background are needed to establish relevant demographic data for school-to-school, district-to-district, and state-to-state comparisons? These and similar questions require value judgements about criteria which include but also transcend technical information.

Balance

A number of policy issues involving the establishment of balance in the statewide assessment and evaluation program have been touched on obliquely above. For example, the number, frequency, and length of tests to be given either on a sampling basis or an every-student basis can be determined at the outer limits by technical considerations--below this minimum, we can say with considerable assurance, the results would be statistically suspect; beyond this limit we would be indulging in overkill. But within the technically-established limits are a host of questions to be decided at the policy level.

Some of the pioneering states--who are certainly to be applauded for their forward-looking efforts--have run into serious problems in certain programs for assessing the results of school improvement programs by means of extensive testing and detailed reporting. The resentment and backlash from the teachers has been paralleled by complaints from parents and students that all the students are doing is taking tests. Some programs which would appear to be of great potential value have had to be suspended or modified. New policies which specify the desired balance between testing and teaching appear to be needed.

There are, certainly, limits imposed by society and parental sensibilities regarding the content of testing. We are not speaking here about what has been called the "Hatchflap"--extreme and seemingly unreasonable objections to testing (and curriculum content) which invoke the Hatch Amendment but go far beyond the intent of the author of that legislation. Rather, most reasonable and thoughtful parents understandably object to overly-intrusive questioning of their children's beliefs and values; this concern is echoed by most educators. Well-thought-out and clearly-articulated policies coming from the state level should be able to set forth a balance between over-cautious and unnecessarily intrusive testing practices.

The sheer amount of data being collected is apparently becoming a problem in some states. Reference has already been made to the difficulties which may emerge when sufficient distinction has not been made between the nice-to-know and the need-to-know indicators and other data items. What seems to be indicated is not, of course, a specific-number-of-items policy, but a clear policy statement which recognizes the limits that must be imposed on all of the elements of a statewide assessment program, and a commitment to maintaining balance throughout the entire undertaking.

Autonomy

There are some potentially serious issues which seem to be emerging about the relationships which should ideally exist among and between local, state, and federal interests in education. As assessment and evaluation plans are devised to encompass an increasingly comprehensive look at educational programs and an increasingly complex analysis of

educational outcomes, the relative importance which should be attached to the interests and concerns of the various levels of the educational structure warrants some thoughtful policy consideration.

State-level decision makers are inevitably the ones caught in the middle. To assure the adequate collection, interpretation, and dissemination of data about the American educational system, the groups working at the national level, both federal and nonfederal, have needs for certain kinds of data at a specified level of uniformity to assure that useful information will be available and a basis for fair comparisons will be established. For the same reasons, the states need comprehensive and uniform data from the local districts. But at every level of the system, the perceptions of needs, obligations, and rights differ.

Confrontational policies which seek jealously (or at least zealously) to protect "turf" or to assert "rights" have rarely been productive. Arguments about "control"—local, state, or federal--generally fizzle out into inconclusive mumbling.

Some tentative principles might be enunciated from which policies appropriate to each level can be formulated. The first is that there <u>is</u> an inherent conflict of interests and perceptions which must be openly recognized. There is no use kidding ourselves that we in education are all one happy family, pursuing universally accepted educational goals. There are at the various educational levels legitimate differences of opinion and different needs. A corollary to the acceptance of legitimate differences in points of view is acceptance of the necessity for compromise: everybody is going to have to give a little.

Once these principles are accepted, some roads to solution become clearer. In the framework of accepting differences and encouraging compromise, programs which recognize a reasonable degree of autonomy for each level involved can be worked out if the process is cooperative and collegial. The apparent success of the CCSSO Assessment Center in reconciling many of the differences among the states is encouraging.

In all candor, though, collegiality alone won't get the job done. State responsibilities may sometimes require a degree of firmness or even stubbornness--the state must decide what data it simply has to require of the local districts, and what it can and cannot provide to national authorities and organizations. Likewise, a local district may have the obligation to staunchly resist, insofar as it is legally possible, serious encroachments on the resources and programs of the district.

Perhaps the overriding principle from which policy in this area may be derived is one which might be called minimalism: keep the entire statewide assessment program (and the corollary programs at the national level) as simple as possible--the smallest number of assessment instruments, the least complex reporting requirements, the least intrusion into normal operating procedures, the minimum of threats to institutional autonomy at every level.

Specific policy options, then, can be developed out of the peculiar circumstances in each state, and from among these options may be chosen concrete policies appropriate to that state.

Pervasive Societal Concerns

It seems unlikely that any statewide assessment and evaluation program can prudently be formulated, implemented, or amended without some policy guidelines which reflect the education decision-makers' best

judgement about the broad social implications of the policy. One need only to observe the reported problems emerging from a number of the states which have embarked on large-scale testing programs to see some of the perils and pitfalls involved.

A case in point would be the concern that is being expressed about the effect of overemphasis in student-achievement testing programs on rather simple cognitive learnings to the detriment of concern with higher-order thinking skills and learnings in the affective domain. To some extent, at least, the nature of the testing and evaluation instruments and programs determine where the instructional and curricular emphasis will lie. Educational program options are sometimes dominated (or circumscribed) by testing practices. For example, a comprehensive school improvement program may be diminished in its effectiveness if only a limited range of educational outcomes continues to be tested.

Another issue which may have been insufficiently addressed in formulating state programs which seek to judge the effectiveness of the educational enterprise is the problem of failure. When essentially inflexible standards are set, and the tests measure to what degree these standards have been met, there will be some who fail--fail to be promoted, fail to graduate, or in the case of teachers, fail to gain admission to teacher education programs, to qualify for certification, or to be eligible to keep the jobs for which they had previously qualified under standards in force at an earlier time.

Failure is, of course, one of the inevitable outcomes of any program of standard-setting, but policies need to be in place early in the game which will minimize both the chance of failure and its devastating effects. Longer lead times before the tests are instituted; ample

provision for re-examination; specific and detailed programs for remediation; and some provision for exceptions in exceptional circumstances--all of these might well become elements of the overall policy for minimizing the traumas often associated with statewide assessment and evaluation programs.

Another of the insufficiently examined problems would seem to be the pervasive social dilemma of how to test and report in a fair and appropriate way without further exacerbating problems of socioeconomic and racial equity. To put it bluntly, with the instruments we now use we can almost guarantee that those who are tested will have scores which place the poor and minority students (and often, minority teachers) at the lower end of the scale. To some extent the same phenomenon applies to individual schools, districts, or states: those with fewer resources are likely to "look bad" regardless of the efforts they expend or the progress they make toward their goals.

The seriously divisive consequences of some testing programs may well call for policies which represent hard choices to be made: shall we just let the chips fall ("that's just the way it <u>is</u>"), or shall we ease up so that the affected groups are not "disproportionately represented"? Middleground policies, which might include some of the mitigating actions suggested above in connection with the problem of failure, may well provide the basis for actions which avoid either extremes: hurting cruelly or dishonestly juggling the standards.

In Conclusion

In examining the problems that complicate statewide assessment and evaluation programs of all sorts, it becomes evident that it is much easier to pinpoint the issues than to develop a full spectrum of mutually exclusive policy options which would assure us that this or this or this could be done with fairly predictable results following in each case. The options are not that discrete; they inevitably overlap. But in every case, with each issue explored, it is clear which there are different directions that may be chosen and that each direction will have fairly foreseeable consequences. Making these choices and accepting the consequences remains the primary function of state education policy makers.

Yet to say that such policy decisions are the ultimate responsibility of state-level decision-makers is not to suggest that these officials are free to make any decision they please if their intent is to employ assessment and evaluation for going beyond such traditional purposes as (1) demonstrating accountability, (2) reporting to the public, and (3) making comparisons among and between schools, districts, or state school systems. To transcend these commendable uses of assessment and evaluation programs into relying on the data obtained in order to determine whethger the schools have succeeded in bringing about specific organizational changes, modifying administrative practices, and improving curriculum and instruction to raise educational standards and increase the level of student achievement--this is ultimately what is required to go "beyond the wall chart."